A METHOD OF DETERMINING ESTIMABLE FUNCTIONS AND TESTABLE HYPOTHESES IN EXPERIMENTAL DESIGN

John Earl Johnson

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

A METHOD OF DETERMINING ESTIMABLE FUNCTIONS
AND
TESTABLE HYPOTHESES IN EXPERIMENTAL DESIGN

by

John Earl Johnson

March 1976

Thesis Advisor: D. R. Barr

Approved for public release; distribution unlimited.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Method of Determining Estimable Functions and Testable Hypotheses in Experimental Design		5. TYPE OF REPORT & PERIOD COVERED Master's Thesis March 1976
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) John Earl Johnson		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		12. REPORT DATE March 1976
		13. NUMBER OF PAGES 39
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Naval Postgraduate School Monterey, California 93940		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

** Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Applications of the general linear model in experimental design and analysis usually involve design matrices of less than full column rank. This may present a problem in determining what elements and functions of the parameter vector are estimable and what hypotheses are testable. This thesis discusses two methods of answering questions about estimability and testability, where the form of the design matrix

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
(Page 1) S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

determines the method to be used. The two methods, both of which can use computer routines, are: (1) direct mathematical computational approach, and (2) a modification of an analysis of variance routine, with a special case of this method using a modified ANOVA routine and solutions to systems of linear equations. Confounding of effects is developed mathematically in connection with determining estimable functions. Methods discussed in this thesis can be applied to the area of Army Test and Evaluation.

A Method of Determining Estimable Functions
and
Testable Hypotheses in Experimental Design

by

John Earl Johnson
Captain, United States Army
B.S., University of Southern Mississippi, 1967

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
March 1976

ABSTRACT

Applications of the general linear model in experimental design and analysis usually involve design matrices of less than full column rank. This may present a problem in determining what elements and functions of the parameter vector are estimable and what hypotheses are testable. This thesis discusses two methods of answering questions about estimability and testability, where the form of the design matrix determines the method to be used. The two methods, both of which can use computer routines, are: (1) direct mathematical computational approach, and (2) a modification of an analysis of variance routine, with a special case of this method using a modified ANOVA routine and solutions to systems of linear equations. Confounding of effects is developed mathematically in connection with determining estimable functions. Methods discussed in this thesis can be applied to the area of Army Test and Evaluation.

TABLE OF CONTENTS

LIST OF FIGURES

I. INTRODUCTION

The model of interest is the general linear model:

$$\tilde{y} = X\tilde{b} + \tilde{e}$$

where $\tilde{y}$ is an nx1 vector of observations,

X is an nxp matrix of known values,

$\tilde{b}$ is a px1 vector of parameters,

and $\tilde{e}$ is a vector of random error terms.

Throughout this thesis a capital letter will denote a matrix, with a prime or superscript of "-1" representing its transpose or inverse respectively. A lower case letter with a tilda (~) above it will denote a vector with a prime again representing its transpose. This model is used in regression analysis and design analysis [Ref. 10]. The derivation of a least squares estimator for $\tilde{b}$ involves minimizing the sum of squares of the difference between the observations and their expected values. This problem can be further reduced to the problem of solving the normal equations

$$X'X\hat{\tilde{b}} = X'\tilde{y}, \tag{1}$$

in which $\hat{\tilde{b}}$ is the estimator of $\tilde{b}$. In the case where X is of full rank p, a unique solution to the normal equations exists. A solution to the normal equations may then be written in form

$$\hat{\tilde{b}} = (X'X)^{-1}X'\tilde{y} \tag{2}$$

The solution $\hat{\tilde{b}}$ found under these conditions can be shown to be the best linear unbiased estimator (b.l.u.e.) of $\tilde{b}$. References 4, 9, and 10 contain proofs of these statements.

If the model is not of full rank, then $(X'X)^{-1}$ does not exist and a solution to the normal equations may be written in terms of a generalized inverse of $X'X$. A generalized inverse G of a matrix A is defined to be any matrix having the property that $AGA = A$. The linear model with a matrix X that is not of full column rank can arise in various ways. For example, the nature of the experiment may result in a design matrix X that is not of full column rank, or the experimenter may have had problems that caused the experiment not to be conducted in accordance with an original full rank design. In this case an analyst may wish to perform a "salvage operation" to gain as much information as possible from the data derived from the experiment. In any case, the analyst is faced with these questions:

(1) What is the actual form of the experiment conducted (i.e., what is X)?

(2) What inferences can be made from the information attained (i.e., what effects (elements and functions of $\tilde{b}$) are estimable and what hypotheses are testable)?

With the less than full rank model, the solution to the normal equations is not unique; rather, many solutions exist. Based on a generalized inverse G of $X'X$, a solution to (2) may be written in the form

$$\tilde{b}^o = GX'\tilde{y} \qquad (3)$$

For each generalized inverse G, discussed in detail by Pringle and Rayner [Ref. 8], there is a solution $\tilde{b}^o$ given by (3), and conversely (i.e., all solutions can be expressed in the form of (3)). Searle [Ref. 10] states that the normal equations are consistent and that the solutions to (1) are given by (3) if and only if G is a generalized inverse of X'X.

This thesis develops and discusses several methods for an experimenter to determine the estimability and testability of linear functions of the parameter vector. First, a mathematically straightforward solution to these problems will be outlined. Then, the use of an analysis of variance computer routine for implementation of this solution is discussed.

II. DEVELOPMENT

To answer questions about testability, several topics must be discussed, including the generalized inverse of $X'X$ and estimability of a linear function of $\tilde{b}$.

A. GENERALIZED INVERSE MATRIX

An introduction to the theory of generalized inverses is contained in Searle [Ref. 10]. A detailed discussion is available in Pringle and Rayner [Ref. 8]. Throughout this paper the symbol G will be used to represent a generalized inverse of $X'X$. As discussed in Reference 8, the matrix G has many alternate names such as "pseudo-inverse," "conditional inverse" and "g-inverse," which makes identical information available in the literature under several different names. Searle presents the following important properties of a generalized inverse:

<u>Theorem 1</u>. When G is a generalized inverse of the matrix $X'X$, then

1. G' is also a generalized inverse of $X'X$;
2. $XGX'X = X$; GX' is a generalized inverse of X;
3. XGX' is invariant to G.

Various methods for computing a generalized inverse are described in detail by Searle and by Pringle and Rayner. The properties of G stated in Theorem 1 are important for the derivations in this thesis, but the direct computation of G will not be required for reasons described in the next chapter.

B. EXPECTED VALUES AND THE MATRIX H

Since $X'X$ is, in general, not of full rank, and thus equation (2) cannot be solved for a unique solution $\hat{\tilde{b}} = (X'X)^{-1}X'\tilde{y}$ as in the full rank model, the normal equations for the less than full rank model are written as

$$X'X\tilde{b}^{o} = X'\tilde{y}$$

in which $\tilde{b}^{o}$ denotes any one of the many solutions that exist. Letting G denote a generalized inverse matrix of $X'X$, then the corresponding solution is given by

$$\tilde{b}^{o} = GX'\tilde{y} \qquad (4)$$

The expected value of $\tilde{b}^{o}$ is given by

$$E(\tilde{b}^{o}) = GX'E(\tilde{y}) = GX'X\tilde{b} = H\tilde{b};$$

where $H = GX'X$. According to Searle [Ref. 10], the matrix H is unique although G is not. For this case, note that $\tilde{b}^{o}$ is an unbiased estimator of $H\tilde{b}$, rather than of $\tilde{b}$.

C. ESTIMABLE FUNCTIONS

Searle [Ref. 10] states that a linear function of parameters, in this case of components of $\tilde{b}$, is estimable if it is identically equal to some linear function of the expected value of $\tilde{y}$, the vector of observations. In other words, $\tilde{q}'\tilde{b}$ is estimable if and only if $\tilde{q}'\tilde{b} = \tilde{t}'E(\tilde{y})$ for some vector $\tilde{t}'$. The vector $\tilde{t}'$ may not be unique.

There are several important properties of estimable functions that will be necessary in determining which hypotheses are testable (see Reference 10).

1. Linear combinations of estimable functions are estimable.

2. All estimable functions are linear combinations of $X\tilde{b}$. The expected value of $\tilde{y}$, $E(\tilde{y})$, is equal to $X\tilde{b}$. By definition, if $\tilde{q}'\tilde{b}$ is estimable, then $\tilde{q}'\tilde{b} = \tilde{t}'E(\tilde{y})$ for some $\tilde{t}'$, so

$$\tilde{q}'\tilde{b} = \tilde{t}'X\tilde{b}. \tag{5}$$

The concept of estimability does not depend on the value $\tilde{b}$, so equation (5) must be true for all values of $\tilde{b}$. Therefore

$$\tilde{q}' = \tilde{t}'X$$

for some vector $\tilde{t}'$. We thus arrive at the following important characterization:

$\tilde{q}'\tilde{b}$ is estimable if and only if $\tilde{q}' = \tilde{t}'X$ for some $\tilde{t}'$.

3. When $\tilde{q}'\tilde{b}$ is estimable, $\tilde{q}'\tilde{b}^o$ is invariant to whatever solution $\tilde{b}^o$ of the normal equations is used. This is true because, by the previous property, for some $\tilde{t}'$,

$$\tilde{q}'\tilde{b}^o = \tilde{t}'X\tilde{b}^o = \tilde{t}'XGX'\tilde{y}$$

and XGX' is invariant under selections of G (Theorem 1). Thus, $\tilde{q}'\tilde{b}^o$ is invariant under choices of G and hence to $\tilde{b}^o$, when $\tilde{q}'\tilde{b}$ is estimable.

4. The following theorem [Ref. 10] provides a procedure for checking the estimability of $\tilde{q}'\tilde{b}$.

<u>Theorem 2.</u> A given function $\tilde{q}'\tilde{b}$ is estimable if and only if $\tilde{q}'H = \tilde{q}'$.

Proof: If $\tilde{q}'\tilde{b}$ is estimable, then the definition of estimability implies that $\tilde{q}' = \tilde{t}'X$ for some $\tilde{t}'$, and $\tilde{q}'H = \tilde{t}'XH = \tilde{t}'XGX'X = \tilde{t}'X$ by Theorem 1. If $\tilde{q}'H = \tilde{q}'$, then $\tilde{q}' = \tilde{q}'GX'X = \tilde{t}'X$ for $\tilde{t}' = \tilde{q}'GX$.

D. TESTABLE HYPOTHESIS

A testable hypothesis is a hypothesis that can be expressed in terms of estimable functions. Assume that a null hypothesis takes the form

$$H_o\colon \tilde{q}'\tilde{b} = 0. \qquad (6)$$

If the null hypothesis $\tilde{q}'\tilde{b} = 0$ is to be tested by the analyst, then $\tilde{q}'\tilde{b}^o$ will be part of the test statistic which will need to be invariant to $\tilde{b}^o$ (detailed proof contained in Ref. 10). As discussed earlier, $\tilde{q}'\tilde{b}^o$ is invariant if $\tilde{q}'\tilde{b}$ is estimable. Thus by applying Theorem 2, if $\tilde{q}'H = \tilde{q}'$ then $\tilde{q}'\tilde{b}$ is estimable and $H_o\colon \tilde{q}'\tilde{b} = 0$ is a testable hypothesis. Similarly, if $\tilde{q}'H \neq \tilde{q}'$, then $\tilde{q}'\tilde{b}$ is not estimable, and therefore the hypothesis $\tilde{q}'\tilde{b}$ is not a testable hypothesis. In the testing of a hypothesis, Searle [Ref. 10] proves that if an analyst uses the standard analysis of variance procedure to "test" the hypothesis given by equation (6) when in fact $\tilde{q}'\tilde{b}$ is not estimable, the actual hypothesis tested is

$$H_o': \tilde{q}'H\tilde{b} = 0$$

The discussions concerning estimability of $\tilde{q}'\tilde{b}$ and the testing of the hypothesis $H_o: \tilde{q}'\tilde{b} = 0$ also can be applied in the form of $Q'\tilde{b}$ for

$$Q'\tilde{b} = \{\tilde{q}_i'\tilde{b}\} \quad \text{for } i = 1, \ldots, s,$$

where $\tilde{b}$ is a px1 vector, $\tilde{q}_i'$ is 1xp, and Q' is a sxp matrix. Thus $Q'\tilde{b}$ is estimable if and only if $Q'H = Q'$, and the hypothesis H_o: $Q'\tilde{b} = 0$ is testable if $Q'\tilde{b}$ is estimable.

The next chapter describes methods of computing H, and the use of the matrix H will be demonstrated in examples in Chapter IV.

III. METHODS OF DETECTING THE MATRIX H

Determination of the matrix H is useful in answering questions concerning estimable functions and testable hypotheses. As discussed in Chapter II, the function $\tilde{q}'\tilde{b}$ is estimable if and only if $\tilde{q}'H = \tilde{q}'$, and the hypothesis H_0: $\tilde{q}'\tilde{b} = 0$ is testable only if $\tilde{q}'\tilde{b}$ is estimable. Recalling that $H = GX'X$ where G is a generalized inverse of the matrix $X'X$, the problem indirectly becomes one of computing the matrix G and then performing the matrix multiplication to obtain $GX'X$. For many designs encountered in practice, the matrices X, X', and G are of large dimensions. Two approaches for computing H will be discussed and demonstrated by examples. The first approach will be the straightforward mathematical approach; but the most practical method appears to be one utilizing an analysis of variance (ANOVA) computer program. The approach using ANOVA computer programs is practical because the analyst would presumably be using a program of that type to analyze test data anyway. As will be discussed, with a simple modification of program inputs, the matrix H can be computed using the ANOVA program.

A. DIRECT MATHEMATICAL COMPUTATION OF MATRIX H

Although the direct approach is demonstrated by a sample problem with a relatively small dimensional design matrix X, this approach provides insights into determining estimability.

For ease of discussion, the sample problem will be presented as a step-by-step procedure.

For this example, the model can be written in the form

$$y_{ij} = m + a_i + e_{ij}, \qquad \text{for } i = 1,\ 2, \text{ and } j = 1,\ 2.$$

In matrix form this becomes

$$\begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} m \\ a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} e_{11} \\ e_{12} \\ e_{21} \\ e_{22} \end{pmatrix}$$

1. <u>Compute X'X</u>

As described in the Chapter II, the transpose product of X is desired throughout the computations.

$$X'X = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{pmatrix}$$

Note that the rank of X'X is 2.

2. <u>Compute G</u>

Several methods are available to describe the computation of the generalized inverse of a matrix [Refs. 8 and 10]. The method in this example is described by Searle [Ref. 10]. The order of X'X is 3 and its rank is 2, so

the first step is to delete (3-2) rows and corresponding columns from $X'X$, to leave a sub-matrix of full rank 2 called $(X'X)_n$. For this example the first row and column are deleted to yield

$$(X'X)_n = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

$$\text{and } (X'X)_n^{-1} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$$

Thus G, a generalized inverse of $X'X$, is determined by replacing all elements of $(X'X)_n$ by those of its inverse and putting in zero for all other elements of $X'X$. Applying this procedure yields

$$G = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix}.$$

As a check to see if G is a generalized inverse of $X'X$, determine if $X'XGX'X = X'X$:

$$X'XGX'X = \begin{pmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{pmatrix}$$

$$= \begin{pmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{pmatrix} = X'X.$$

3. Compute H

The computation of H is a direct matrix multiplication.

$$H = GX'X = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \qquad (7)$$

4. Compute H Using Alternate Method

Two equations are useful in demonstrating an alternate method for computing H:

$$H = GX'X$$

$$\tilde{b}^o = GX'\tilde{y}, \quad \text{for any } \tilde{y}.$$

If the $\tilde{y}$ used for computation is $\tilde{x}_j$ where $\tilde{x}_j$ is the j^{th} column of the design matrix X, then a solution $\tilde{b}^o_j$ can be computed for each j. Partitioning the matrices X and H, the following result is noted:

$$H = (\tilde{b}_1{}^o \mid \tilde{b}_2{}^o \mid \ldots \mid \tilde{b}_p{}^o) = GX'(\tilde{x}_1 \mid \tilde{x}_2 \mid \ldots \mid \tilde{x}_p) = GX'X.$$

For this thesis the columns of the matrix X, $\tilde{x}_j$, will be referred to as "pseudo-data."

To demonstrate this procedure of computing H, the previous example will be used. The solution becomes one of computing

$$\tilde{b}_j^{\,o} = GX'\tilde{x}_j \qquad \text{for } j = 1, 2, 3. \tag{8}$$

$$\tilde{b}_1^{\,o} = GX'\tilde{x}_1$$

$$= \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} .$$

similarly,

$$\tilde{b}_2^{\,o} = GX'\tilde{x}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} .$$

and

$$\tilde{b}_3^{\,o} = GX'\tilde{x}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} .$$

$$\text{Thus } H = (\tilde{b}_1^{\,o} \mid \tilde{b}_2^{\,o} \mid \tilde{b}_3^{\,o}) = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \text{, as before.} \tag{9}$$

Of course H computed using this method is the same as the H determined in paragraph 3. This method of computing H is important because an ANOVA computer routine can be used to solve for the columns of H, using pseudo-data in place of $\tilde{y}$.

B. USE OF ANOVA ROUTINE TO COMPUTE THE MATRIX H

In section A two methods were discussed for computing H. The simplicity of the design matrix X used in the example permitted easy computations. A more realistic design matrix in practice would be of large dimensions that such hand computations might become impractical. Computers provide a means of solving this problem. One approach on a computer would be a direct mathematical approach using available computer routines for computing G, then H. Three routines would be required to compute H: one routine to compute $X'X$; a second routine to compute the generalized inverse of the matrix $X'X$; and a third routine to compute the matrix product of $GX'X$ to attain H. This method is straightforward but requires detailed user knowledge by the analyst. The limitations of this approach are based on the limitations of the computer routines used, which normally are in the form of dimension capacities.

A second approach using an ANOVA computer routine may be more practical for an analyst, since an ANOVA routine will probably be used to analyze the data from the experiment anyway. The procedure to be used involves the use of pseudo-data as described in step 4 of section A. Pseudo-data is analyzed by the routine, with the results of interest being the regression coefficients as determined by solving the normal equations (equation (1)) under the full model hypothesis.

This procedure can be demonstrated by an example using the General-Linear Hypothesis routine, BMD05V, of the BMD Biomedical Computer Programs [Ref. 11]. Suppose that the model is

$$y_{ij} = m + a_i + c_j + ac_{ij} + e_{ij}, \qquad \text{for } i = 1, 2;$$

$$j = 1, 2, 3.$$

In matrix form this model can be written in the general form

$$\begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} m \\ a_1 \\ a_2 \\ c_1 \\ c_2 \\ c_3 \\ ac_{11} \\ ac_{12} \\ ac_{13} \\ ac_{21} \\ ac_{22} \\ ac_{23} \end{pmatrix} + \begin{pmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{21} \\ e_{22} \\ e_{23} \end{pmatrix} \quad (10)$$

Each column of the design matrix is entered as pseudo-data. The computer output lists estimates of coefficients corresponding to the full model hypothesis. This gives $\tilde{b}_j^{\,o}$ for the j^{th} column of pseudo-data. Using the smaller dimensional

notation required by the BMDO5V routine causes the coefficient vector to be in a reduced form. This can be expanded to the full $\tilde{b}_j^o$ by applying the "usual linear restrictions" on the parameters given by

$$\sum_i a_i = 0$$

$$\sum_j c_j = 0$$

$$\sum_i ac_{ij} = 0, \text{ etc.}$$

These linear restrictions are common restrictions placed on the parameters in ANOVA problems, since by using these restrictions, the "reduced" design matrix may become of full rank, which in turn aides in solving the normal equations. The matrix H can then be formed as previously discussed by taking

$$H = (\tilde{b}_1^o \mid \ldots \mid \tilde{b}_{12}^o).$$

Using the design matrix from equation (10) and the BMDO5V routine, this procedure was followed to determine:

$$H = \begin{pmatrix} 1 & 1/2 & 1/2 & 1/3 & 1/3 & 1/3 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 0 & 1/2 & -1/2 & 0 & 0 & 0 & 1/6 & 1/6 & 1/6 & -1/6 & -1/6 & -1/6 \\ 0 & -1/2 & 1/2 & 0 & 0 & 0 & -1/6 & -1/6 & -1/6 & 1/6 & 1/6 & 1/6 \\ 0 & 0 & 0 & 2/3 & -1/3 & -1/3 & 1/3 & -1/6 & -1/6 & 1/3 & -1/6 & -1/6 \\ 0 & 0 & 0 & -1/3 & 2/3 & -1/3 & -1/6 & 1/3 & -1/6 & -1/6 & 1/3 & -1/6 \\ 0 & 0 & 0 & -1/3 & -1/3 & 2/3 & -1/6 & -1/6 & 1/3 & -1/6 & -1/6 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & -1/6 & -1/6 & -1/3 & 1/6 & 1/6 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1/6 & 1/3 & -1/6 & 1/6 & -1/3 & 1/6 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1/6 & -1/6 & 1/3 & 1/6 & 1/6 & -1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1/3 & 1/6 & 1/6 & 1/3 & -1/6 & -1/6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/6 & -1/3 & 1/6 & -1/3 & 1/3 & -1/6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/6 & 1/6 & -1/3 & -1/6 & -1/6 & 1/3 \end{pmatrix} \qquad (11$$

As a check, this matrix H satisfies the requirement that $X'XGX'X = X'XH = X'X$.

If the transpose product of the design matrix used as input for the BMD05V routine is singular, then the BMD05V routine will terminate with an error statement. In such a case, the design matrix must be modified so as to accomplish the required results.

To outline the procedure in such a case, consider the model

$$y_{ijk} = m + a_i + c_j + d_k + acd_{ijk} + cd_{jk} + ad_{ik} + ac_{ij} + e_{ijk}, \quad \text{for } i = 1, 2;\ j = 1, 2;\ k = 1,2.$$

This model is a 2^3 factorial design, and for this example there will be observations in only four of the eight cells ($\frac{1}{2}$ replication). Suppose observations are numbered and placed as shown below:

	c_1	c_2
a_1	1	
a_2		4

d_1

	c_1	c_2
a_1		2
a_2	3	

d_2

The cells containing numbers represent the locations of the observations. The transpose of the design matrix X and the parameter vector $\tilde{b}$ for this model are shown in figures 1 and 2. The first step in computing H will be to determine the form of the reduced design matrix Z, which would be required by the BMD05V routine. Thus the form would be

$$Z = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix} .$$

This design matrix Z will not run on BMD05V because the product Z'Z is singular. To reduce Z to a form acceptable to the computer routine, determine a set of linearly independent columns of Z and write these columns as the matrix A. For this example, the first four columns suffice:

$$X' = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

FIGURE 1. Transpose of Design Matrix for $\frac{1}{2}$ Replication of a 2^3 Factorial Design.

$$
\tilde{b} = \begin{pmatrix} m \\ a_1 \\ a_2 \\ c_1 \\ c_2 \\ d_1 \\ d_2 \\ acd_{111} \\ acd_{112} \\ acd_{121} \\ acd_{122} \\ acd_{211} \\ acd_{212} \\ acd_{221} \\ acd_{222} \\ cd_{11} \\ cd_{12} \\ cd_{21} \\ cd_{22} \\ ad_{11} \\ ad_{12} \\ ad_{21} \\ ad_{22} \\ ac_{11} \\ ac_{12} \\ ac_{21} \\ ac_{22} \end{pmatrix}
$$

FIGURE 2. Parameter Vector for $\frac{1}{2}$ Replication of a 2^3 Factorial Design.

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

Note that Z is of the form $Z = (A \mid A)$. Since A is nonsingular, so is $A'A$, and the inverse of $A'A$ exists. Let $G_A = (A'A)^{-1}$ and $H_A = G_A\ (A'A) = (A'A)^{-1}(A'A) = I$. The remaining steps are to compute G_Z, the generalized inverse of $Z'Z$ and then to compute H_Z. Observe that

$$Z'Z = \begin{pmatrix} A' \\ A' \end{pmatrix} \quad (A \quad A) = \left(\begin{array}{c|c} A'A & A'A \\ \hline A'A & A'A \end{array}\right) .$$

Using the fact that $A'A\ G_A\ A'A = A'A$, then G_Z is of the form

$$G_Z = \tfrac{1}{4} \left(\begin{array}{c|c} G_A & G_A \\ \hline G_A & G_A \end{array}\right) .$$

The matrix G_Z possesses the property that $Z'Z\ G_Z\ Z'Z = Z'Z$, so,

$$H_Z = G_Z\ Z'Z = \tfrac{1}{4} \left(\begin{array}{c|c} G_A & G_A \\ \hline G_A & G_A \end{array}\right) \left(\begin{array}{c|c} A'A & A'A \\ \hline A'A & A'A \end{array}\right)$$

$$= \tfrac{1}{2} \left(\begin{array}{c|c} G_A A'A & G_A A'A \\ \hline G_A A'A & G_A A'A \end{array}\right) = \tfrac{1}{2} \left(\begin{array}{c|c} I & I \\ \hline I & I \end{array}\right)$$

where I is the 4x4 identity matrix. This H_Z is in a reduced form and thus must be expanded from its 8x8 dimension to the 27x27 H matrix. Additional columns can be added (i.e., rows

expanded) based on the fact that each of the remaining columns can be expressed as a linear combination of the matrix H_{RZ}, where H_{RZ} consists of the first four columns of H_Z. The row expansion of H_{RZ} will produce the 8x27 dimensional matrix H_{FZ}. To expand the rows, recall that each column of the design matrix X, denoted by $\tilde{x}_j$, for j = 5, ...27, can be expressed as a linear combination of the matrix A (the first four columns of X). This linear combination can be expressed as

$$A\tilde{v}_j = \tilde{x}_j, \qquad \text{for } j = 5, \ldots, 27.$$

Since A is nonsingular, A^{-1} exists and

$$\tilde{v}_j = A^{-1}\tilde{x}_j, \qquad \text{for } j = 5, \ldots, 27.$$

The same linear relationship exists between each column of H_{FZ} and H_{RZ}. Letting $\tilde{h}_j$ be the j^{th} column of H_{FZ} and $\tilde{x}_j$ be the corresponding column of pseudo-data from the design matrix X (Figure 1), $\tilde{h}_j$ can be computed as

$$\tilde{h}_j = H_{RZ}\tilde{v}_j, \qquad \text{for } j = 5, \ldots, 27 \text{ in this example.}$$

Each column can be expanded (i.e., additional rows added) by applying the same restraints as before, in the form:

$$\sum_i a_i = \sum_j c_j = \sum_k d_k = \sum_i acd_{ijk} = \sum_j acd_{ijk} = 0$$

$$\sum_k acd_{ijk} = \sum_j cd_{jk} = \sum_k cd_{jk} = \sum_i ad_{ik} = 0 \qquad (12)$$

and $$\sum_k ad_{ik} = \sum_i ac_{ij} = \sum_j ac_{ij} = 0.$$

Using this procedure as described, the 27x27 H for the $\frac{1}{2}$ replication of a 2^3 factorial experimental design was obtained. It is shown in Figures 3 and 4 where $H = (H_1 \mid H_2)$.

$$
H_1 = \begin{pmatrix}
\frac{1}{2} & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & \frac{1}{2} & -\frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 \\
0 & \frac{1}{2} & -\frac{1}{2} & 0 & 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 & -\frac{1}{4} & -\frac{1}{4} & 0 \\
0 & -\frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & -\frac{1}{4} & 0 & 0 & -\frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 \\
0 & 0 & 0 & \frac{1}{2} & -\frac{1}{2} & 0 & 0 & \frac{1}{4} & 0 & 0 & -\frac{1}{4} & 0 & \frac{1}{4} & -\frac{1}{4} & 0 \\
0 & 0 & 0 & -\frac{1}{2} & \frac{1}{2} & 0 & 0 & -\frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 & -\frac{1}{4} & \frac{1}{4} & 0 \\
0 & 0 & 0 & 0 & 0 & \frac{1}{2} & -\frac{1}{2} & \frac{1}{4} & 0 & 0 & -\frac{1}{4} & 0 & -\frac{1}{4} & \frac{1}{4} & 0 \\
0 & 0 & 0 & 0 & 0 & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & -\frac{1}{4} & 0 \\
\frac{1}{2} & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & \frac{1}{2} & -\frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 \\
-\frac{1}{2} & 0 & -\frac{1}{2} & 0 & -\frac{1}{2} & 0 & -\frac{1}{2} & \frac{1}{4} & 0 & 0 & -\frac{1}{4} & 0 & -\frac{1}{4} & -\frac{1}{4} & 0 \\
-\frac{1}{2} & 0 & -\frac{1}{2} & 0 & -\frac{1}{2} & 0 & -\frac{1}{2} & \frac{1}{4} & 0 & 0 & -\frac{1}{4} & 0 & -\frac{1}{4} & -\frac{1}{4} & 0 \\
\frac{1}{2} & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & \frac{1}{2} & -\frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 \\
-\frac{1}{2} & 0 & -\frac{1}{2} & 0 & -\frac{1}{2} & 0 & -\frac{1}{2} & \frac{1}{4} & 0 & 0 & -\frac{1}{4} & 0 & -\frac{1}{4} & -\frac{1}{4} & 0 \\
\frac{1}{2} & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & \frac{1}{2} & -\frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 \\
\frac{1}{2} & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & \frac{1}{2} & -\frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 \\
-\frac{1}{2} & 0 & -\frac{1}{2} & 0 & -\frac{1}{2} & 0 & -\frac{1}{2} & \frac{1}{4} & 0 & 0 & -\frac{1}{4} & 0 & -\frac{1}{4} & -\frac{1}{4} & 0 \\
0 & \frac{1}{2} & -\frac{1}{2} & 0 & 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 & -\frac{1}{4} & -\frac{1}{4} & 0 \\
0 & -\frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & -\frac{1}{4} & 0 & 0 & -\frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 \\
0 & -\frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & -\frac{1}{4} & 0 & 0 & -\frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 \\
0 & \frac{1}{2} & -\frac{1}{2} & 0 & 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 & -\frac{1}{4} & -\frac{1}{4} & 0 \\
0 & 0 & 0 & \frac{1}{2} & -\frac{1}{2} & 0 & 0 & \frac{1}{4} & 0 & 0 & -\frac{1}{4} & 0 & \frac{1}{4} & -\frac{1}{4} & 0 \\
0 & 0 & 0 & -\frac{1}{2} & \frac{1}{2} & 0 & 0 & -\frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 & -\frac{1}{4} & \frac{1}{4} & 0 \\
0 & 0 & 0 & -\frac{1}{2} & \frac{1}{2} & 0 & 0 & -\frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 & -\frac{1}{4} & \frac{1}{4} & 0 \\
0 & 0 & 0 & \frac{1}{2} & -\frac{1}{2} & 0 & 0 & \frac{1}{4} & 0 & 0 & -\frac{1}{4} & 0 & \frac{1}{4} & -\frac{1}{4} & 0 \\
0 & 0 & 0 & 0 & 0 & \frac{1}{2} & -\frac{1}{2} & \frac{1}{4} & 0 & 0 & -\frac{1}{4} & 0 & -\frac{1}{4} & \frac{1}{4} & 0 \\
0 & 0 & 0 & 0 & 0 & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & -\frac{1}{4} & 0 \\
0 & 0 & 0 & 0 & 0 & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & -\frac{1}{4} & 0 \\
0 & 0 & 0 & 0 & 0 & \frac{1}{2} & -\frac{1}{2} & \frac{1}{4} & 0 & 0 & -\frac{1}{4} & 0 & -\frac{1}{4} & \frac{1}{4} & 0
\end{pmatrix}
$$

FIGURE 3. The Matrix H_1.

$$
H_2 = \begin{pmatrix}
-\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\
\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\
-\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\
\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\
-\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \\
\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \\
-\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\
-\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\
\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\
\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\
-\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\
\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\
-\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\
-\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\
\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\
\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\
-\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\
-\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\
\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\
\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\
-\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \\
-\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \\
\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\
\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \\
-\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\
-\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\
-\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{4} \\
\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4}
\end{pmatrix}
$$

FIGURE 4. The Matrix H_2.

IV. APPLICATIONS OF THE MATRIX H

The discussion in Chapter II as to the importance of the matrix H in answering an experimenter's questions on estimability and testability can be applied to the three examples in the previous chapter. In examining these examples, the concept of confounding can be mathematically explained and illustrated. Confounding as defined by References 1, 2, 3, 5, 6 and 7 is the designing or arrangement of an experiment in such a manner that certain effects cannot be distinguished from other effects. The references previously cited discuss the methods for intentionally confounding certain effects, normally the higher order interaction terms, with other effects for fractional replications in several different kinds of experimental designs. As discussed in Chapter II, the linear combination of the parameters $\tilde{q}'\tilde{b}$ is estimable if and only if $\tilde{q}'H = \tilde{q}'$. As stated before, a hypothesis is testable if it consists of estimable functions; so if the hypothesis H_o: $\tilde{q}'\tilde{b} = 0$ is "tested" with the standard ANOVA approach, and $\tilde{q}'\tilde{b}$ is not estimable, the hypothesis actually tested is in the form H_o' : $\tilde{q}'H\tilde{b} = 0$. From the expression $\tilde{q}'H\tilde{b}$ the analyst can determine which effects are confounded in the design. This determination of confounding can be illustrated through examples.

In the first example, the matrix H was determined to be

$$H = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} .$$

Consider the hypothesis H_o: $a_1 - a_2 = 0$ which implies $\tilde{q}' = (0 \quad 1 \quad -1)$. For $\tilde{q}'\tilde{b}$ to be estimable and thus H_o: $\tilde{q}'\tilde{b} = 0$ testable, the condition $\tilde{q}'H = \tilde{q}'$ must be met.

$$\tilde{q}'H = (0 \quad 1 \quad -1) \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} = (0 \quad 1 \quad -1) = \tilde{q}'.$$

Therefore, the hypothesis H_o: $a_1 - a_2 = 0$ is testable. Consider a different hypothesis, say H_o:$a_1 = 0$. Then $\tilde{q}' = (0 \quad 1 \quad 0)$. Checking for estimability:

$$\tilde{q}'H = (0 \quad 1 \quad 0) \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} = (1 \quad 1 \quad 1) \neq \tilde{q}'.$$

Therefore, the H_o:$a_1 = 0$ is not testable but the hypothesis tested actually is of the form H_o: $\tilde{q}'H\tilde{b} = 0$ or H_o: $m + a_1 = 0$. This result means that the a_1 effect is confounded with the mean.

In the second example, H was given by equation (1). Checking the testability of the hypothesis H_o: $a_1 - a_2 = 0$ where

$$\tilde{q}' = (0 \quad 1 \quad -1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0),$$

$\tilde{q}'H$ does equal $\tilde{q}'$; therefore, the hypothesis H_o: $\tilde{q}'\tilde{b} = 0$ is testable. Suppose that the hypothesis H_o: $c_1 - c_2 = 0$, where $\tilde{q}' = (0\ \ 0\ \ 0\ \ 1\ \ -1\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0)$, is tested. In this case $\tilde{q}'H$ is not equal to $\tilde{q}'$, and therefore the hypothesis is not testable. If the hypothesis H_o: $c_1 - c_2 = 0$ is tested using standard ANOVA test statistics, the actual hypothesis tested would be:

$$H_o' : \tilde{q}'H\tilde{b} = (0\ \ 0\ \ 0\ \ 1\ \ -1\ \ 0\ \ \tfrac{1}{2}\ \ -\tfrac{1}{2}\ \ 0\ \ \tfrac{1}{2}\ \ -\tfrac{1}{2}\ \ 0)\ \tilde{b} =$$

$$c_1 - c_2 + \tfrac{1}{2}ac_{11} - \tfrac{1}{2}ac_{12} + \tfrac{1}{2}ac_{21} - \tfrac{1}{2}ac_{22} = 0$$

Confounding obviously is present in the form of the two way interaction terms ac_{ij}. The analyst may suspect that the interaction terms are not significant, in which case they can be disregarded. In the present case, if $\sum_i ac_{ij} = 0$ for $j = 1, 2$ is assumed as part of the model, again H_o is testable.

The $\frac{1}{2}$ replication of the 2^3 factorial design produces confounding of many effects which can be illustrated for the case of "testing" the hypothesis H_o: $a_1 - a_2 = 0$. For this example $\tilde{q}'$ is 1×27 and is of the form:

$$\tilde{q}' = (0\ \ 1\ \ -1\ \ 0\ \ .\ .\ .\ \ 0).$$

It is easily verified that $\tilde{q}'H \neq \tilde{q}$, implying that $\tilde{q}'\tilde{b}$ is not estimable, and therefore the hypothesis H_o: $a_1 - a_2 = 0$ is not testable. By computing $\tilde{q}'H\tilde{b}$, the form of the confounding can be obtained:

$$\tilde{q}'H\tilde{b} = a_1 - a_2 + \tfrac{1}{2}\, acd_{111} + \tfrac{1}{2}\, acd_{122} - \tfrac{1}{2}\, acd_{212} - \tfrac{1}{2}\, acd_{221}$$

$$+ \tfrac{1}{2}\, cd_{11} - \tfrac{1}{2}\, cd_{12} - \tfrac{1}{2}\, cd_{21} + \tfrac{1}{2}\, cd_{22} + \tfrac{1}{2}\, ad_{11}$$

$$+ \tfrac{1}{2}\, ad_{12} - \tfrac{1}{2}\, ad_{21} - \tfrac{1}{2}\, ad_{21} - \tfrac{1}{2}\, ad_{22} + \tfrac{1}{2}\, ac_{11}$$

$$+ \tfrac{1}{2}\, ac_{12} - \tfrac{1}{2}\, ac_{22} = 0.$$

V. CONCLUSIONS AND RECOMMENDATIONS

The matrix H, once computed, is valuable to the analyst for determining the estimable linear functions of the parameter vector and for ascertaining which hypotheses are testable. As discussed in the previous chapter, information concerning confounding can be gained through H.

Of particular note is the fact that the computation of the matrix H is dependent on the design matrix X. Although an experiment may not have been conducted in accordance with original design, the design matrix X will be known. The methods of computing H discussed in this thesis have limitations. The direct mathematical approach is limited by the constraints of the computer routines used to compute X'X, the generalized inverse matrix G of X'X, and thus H. The second method using the analysis of variance routine with its dimensionality constraints requires knowledge in the reduction in dimensionality of the design matrix and the row expansion of H, but this knowledge would be required to use the ANOVA routine anyway. The second method is constrained by the fact that the ANOVA routine will not run for a singular "reduced" X'X; this constraint led to a modification of the second method. The limitations of the modified method also would be in the form of computer program constraints in computing matrix inverses, solving systems of linear equations to expand the columns, and expanding the rows using the linear restrictions (equation (12)) imposed by BMD05V.

A special purpose computer program to develop H could be written. A desired computer routine would be one that would accept as input the reduced design matrix and a set of $\tilde{q}$'s and produce as output the corresponding matrix H, and $\tilde{q}$'H's. Although such a computer program might have some of the limitations discussed previously, the program could answer the user's questions concerning estimability and testability. As discussed in the previous chapter, the user could gain information as to which effects were confounded, and this in turn would aid in making decisions as to the significance of this confounding in his experiment.

LIST OF REFERENCES

1. Cochran, W. G. and Cox, G. M., Experimental Designs, Wiley, 1950.

2. Cox, D. R., Planning of Experiments, Wiley, 1958.

3. Davies, O. L., The Design and Analysis of Industrial Experiments, 2nd ed., Hafner, 1956.

4. Graybill, F. A., An Introduction to Linear Statistical Models, v. 1, McGraw-Hill, 1961.

5. Hicks, C. R., Fundamental Concepts in the Design of Experiments, Holt, Rinehart and Winston, 1964.

6. John, P. W. M., Statistical Design and Analysis of Experiments, MacMillan, 1971.

7. Kirk, R. E., Experimental Design: Procedures for the Behavioral Sciences, Brooks/Cole, 1968.

8. Pringle, R. M. and Rayner, A. A., Generalized Inverse Matrices with Applications to Statistics, Hafner, 1971.

9. Scheffé, H., The Analysis of Variance, Wiley, 1959.

10. Searle, S. R., Linear Models, Wiley, 1971.

11. University of California Press, BMD Biomedical Computer Program, 1973.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0212 Naval Postgraduate School Monterey, California 93940	2
3.	Department Chairman, Code 55 Department of Operations Research and Administrative Sciences Naval Postgraduate School Monterey, California 93940	1
4.	Associate Professor Donald R. Barr, Code 55Bn Department of Operations Research and Administrative Sciences Naval Postgraduate School Monterey, California 93940	1
5.	U. S. Army Military Personnel Center ATTN: DAPC-OPD-PD-CS 200 Stovall Street Alexandria, Virginia 22232	1
6.	CPT John E. Johnson 416 Williams Street Hattiesburg, Mississippi 39401	1

Thesis 165102
J6223 Johnson
c.1 A method of determining estimable functions and testable hypotheses in experimental design.